A Tutorial on Adjoint Methods of Calculating Data Kernels

Bill Menke, January 2007

Consider an egg cooking on a frying pan on a stove with a flickering flame. The flame is the source of heat that governs the temperature of the frying pan though the physics of heat transfer. The opacity of the egg white after 3 minutes of cooking is the observable. It depends upon the temperature history of the frying pan. We might ask, "How much is the opacity perturbed, if we were to increase the heat of the flame by 10% during the first minute of cooking?"

This type of question can be addressed with so-called *adjoint methods*. The purpose of this tutorial is to explain this method, and to apply it to the egg example posed above.

1. **What is an adjoint?** The adjoint of a differential operator is analogous to the transpose of a matrix. If you don't immediately find that analogy helpful, then we need to develop it step-by-step. Thus tutorial assumes that you know more about matrices and vectors than you know about continuous functions. If this is the case, read on!

There is an obvious analogy between a vector, $\mathbf{v}$, and a continuous function, a(t). A vector, $\mathbf{v}$, with components $\{\mathbf{v}_i, i=1, \ldots N\}$ has a value for each possible value of its index, i. A function, a(t) has a value for each possible value of its domain, t.

We use this analogy in time-series analysis when we approximate a smooth function by its values at a sequence of N evenly-spaced points, t_i (with spacing Δt). We can write $v_i=a(t_i)$. In fact, you might even go as far as saying that a continuous function is a limiting case of a vector – the limit where $N\rightarrow\infty$ while $\Delta t\rightarrow 0$.

The dot product of two N-dimensional vectors, $\mathbf{u}$ and $\mathbf{v}$, is a scalar quantity defined as:

$$\mathbf{u}\cdot\mathbf{v} \equiv \Sigma_{i=1}^{N} u_i v_i \qquad (1)$$

An analogous operation with continuous functions is the so-called *inner product*. Given two functions, a(t) and b(t), defined on the domain $t_1\leq t\leq t_2$, their inner product is:

$$(a,b) \equiv \int_{t1}^{t2} a(t)\, b(t)\, dt \qquad (2)$$

Like the dot product, the inner product of two functions is a scalar quantity. The relationship between inner products and dot products becomes apparent if we represent the functions a(t) and b(t) by their values at a sequence of N evenly-spaced points, t_i (with spacing Δt). We can write $u_i=a(t_i)$ and $v_i=b(t_i)$ and then approximate the integral in Equation 2 by its Riemann sum, thus obtaining the relationship:

$$(a,b) \approx \Delta t\, \Sigma_{i=1}^{N} a(t_i)\, b(t_i) = \Delta t\, \mathbf{u}\cdot\mathbf{v} \qquad (3)$$

Let's now turn our attention to matrix multiplication. Suppose that we multiply a vector **v**, by a square matrix, **C**. The result is itself a vector:

$$(\mathbf{Cu})_i = \Sigma_{j=1}^{N} C_{ij}\, u_j \qquad (4)$$

In the world of continuous functions, the linear differential operator, *L*, plays a role analogous to a square matrix, **C**. If, a(t) is a function, and *L* is a linear differential operator, then *L*a(t) is a function. The analogy can be understood by considering a simple operator such as *L*=d/dt. Suppose that we represent a(t) and *L*a(t)=da/dt by their values at a sequence of N evenly-spaced points, t_i (with spacing Δt). Call these two vectors **u** and **v**, respectively, that is, $u_i=a(t_i)$ and $v_i=da/dt|_{t_i}$. Then by the definition of differentiation, we can write:

$$da/dt|_{t_i} \approx (1/\Delta t)\,[a(t_i+\Delta t) - a(t_i)] \quad \text{or} \quad v_i = (1/\Delta t)\,[u_{i+1} - u_i] \text{ or } \mathbf{v}=\mathbf{Cu} \qquad (5)$$

with the matrix **C** given by:

$$\mathbf{C} = (1/\Delta t) \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ & & \dots & & 0 \\ \dots & & 0 & -1 & 1 \end{bmatrix} \qquad (6)$$

This is the so-called finite-difference approximation to the derivative. Note that the matrix, **C**, in equation 6 has one fewer row than it has columns, and is thus not square. To make it square, and thus to ensure a unique solution of an equation involving **C**, we must add a row that represents a boundary condition. A boundary condition that specifies a(t) at t_1 (the left hand side of the interval), for example, would be equivalent to adding the row [1, 0, 0, 0, … 0] to the top of **C**.

Above, we have identified *L* as a linear differential operator. Actually, any linear integral operator would work in this context, too. Thus, for instance, if $L\text{a}(t)= \int_0^t a(t')\, dt'$ then the Riemann summation approximation to integration implies:

$$\mathbf{C} = \Delta t \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ & & \dots & & \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \qquad (6)$$

So from now on, we will use the phrase *linear operator* to include both the differential and integral kinds.

Now back to dot products. Suppose we form a new vector, $\mathbf{Cu}$, by multiplying the vector, $\mathbf{u}$, by a square matrix, $\mathbf{C}$. Similarly, we form a new vector, $\mathbf{Dv}$, by multiplying the vector, $\mathbf{v}$, by a square matrix, $\mathbf{D}$. We can now ask a critical question: What must the relationship between $\mathbf{C}$ and $\mathbf{D}$ be so that $(\mathbf{Cu})\cdot\mathbf{v}$ is always equal to $\mathbf{u}\cdot(\mathbf{Dv})$?

The answer is that $\mathbf{C}$ and $\mathbf{D}$ must be matrix transposes of one another. This result can be derived by combining the definition of matrix multiplication:

$$(\mathbf{Cu})_i = \Sigma_{j=1}^{N} C_{ij}\, u_j \quad \text{and} \quad (\mathbf{Dv})_i = \Sigma_{j=1}^{N} D_{ij}\, v_j \tag{7}$$

with the definition of the dot product in Equation 1:

$$(\mathbf{Cu})\cdot\mathbf{v} = \Sigma_{i=1}^{N} \Sigma_{j=1}^{N} C_{ij}\, u_j\, v_i \quad \text{and} \quad \mathbf{u}\cdot(\mathbf{Dv}) = \Sigma_{i=1}^{N} \Sigma_{j=1}^{N} D_{ij}\, u_i\, v_j \tag{8}$$

These two quantities are only equal when $C_{ij}= D_{ji}$, that is, $\mathbf{C}$ is the matrix transpose of $\mathbf{D}$: $\mathbf{C}=\mathbf{D}^T$. Matrix transposes have well understood properties: for instance, if $\mathbf{C}=\mathbf{D}^T$ then $\mathbf{D}=\mathbf{C}^T$; and $\mathbf{D}=(\mathbf{D}^T)^T$; and $(\mathbf{C}^T)^{-1}=(\mathbf{C}^{-1})^T$; and $(\mathbf{C}+\mathbf{D})^T=\mathbf{C}^T+\mathbf{D}^T$ (where $\mathbf{C}^{-1}$ is the matrix inverse of $\mathbf{C}$). Note that in the special case of a symmetric matrix, $\mathbf{C}$, $(\mathbf{Cu})\cdot\mathbf{v}$ is always equal to $\mathbf{u}\cdot(\mathbf{Cv})$, since $\mathbf{C}=\mathbf{C}^T$.

Here's the analogous critical question for the inner product: What must the relationship between two linear operators, L and M, be so that $(L\text{a}, \text{b})$ is always equal to $(\text{a}, M\text{b})$?

When L and M have this special relationship, L is said to be the *adjoint* to M (and M is said to be the adjoint of L). We can denote this relationship symbolically, with M^* meaning the adjoint to M. Thus $(L\text{a}, \text{b})=(\text{a}, M\text{b})$ when $L=M^*$ (and $M=L^*$). Relationships analogous to those enumerated above for matrix transposes hold for adjoints: $L=(L^*)^*$; and $(L^{-1})^*=(L^*)^{-1}$; and $(L+M)^*=L^*+M^*$ holds for adjoints (we will discuss the meaning of L^{-1}, the inverse of the differential operator, L, below). In the special case where a differential operator, L, satisfies $(L\text{a}, \text{b})=(\text{a},L\text{b})$, then L is said to be self-adjoint, that is $L=L^*$. The adjoint of a linear operator is analogous to the transpose of a matrix. A self-adjoint linear operator is analogous to a symmetric matrix.

Note that we have only given a name to this relationship. Nothing said so far tell us anything about how adjoints can be constructed. We will not examine the problem in detail here. But here's a simple special case. We can use integration by parts to show that the adjoint of $L=\text{d}/\text{dt}$ is $L^*=(-\text{d}/\text{dt})$.

$$(\text{da}/\text{dt}, \text{b}) = \int_{t1}^{t2} \text{da}/\text{dt}\ \text{b}\ \text{dt} = \text{ab}|_{t1}^{t2} - \int_{t1}^{t2} \text{a}\ \text{db}/\text{dt} = (\text{a}, -\text{db}/\text{dt}) \tag{9}$$

Here we have assumed that $a(t_1)=0$ or $b(t_1)=0$ and $a(t_2)=0$ or $b(t_2)=0$, that is that a(t) and b(t) satisfy homogeneous boundary conditions. The operator $L=\text{d}^2/\text{dt}^2$ can easily shown

to be self-adjoint using the same technique. Its also easy to show that the trivial operator, L=c (where c is a constant) is self-adjoint.

2. What is a data kernel? A data kernel relates, via an inner product, a perturbation in some observable, or *datum*, to a perturbation in some parameter affecting the system.

We assume that the underlying dynamics of the system are represented by a linear differential equation relating some scalar field, u(t), to some forcing, f(t):

$$L\, u(t) = f(t) \qquad (10)$$

Here L is a linear differential operator. We also assume that the scalar field, u(t), satisfies known initial conditions. Note that we can write the solution of this equation as:

$$u(t) = \int_{t1}^{t2} F(t,t')\, f(t')\, dt' \equiv L^{-1} f(t) \qquad (11)$$

Here F(t,t') is the Green function of equation 10, that is, the function that solves equation 10 in the special case of an impulsive forcing, $f(t)=\delta(t-t')$ (where δ is the Dirac impulse function). We symbolically write integration of the Green function as the linear operator, L^{-1}, that is, in an abstract sense L^{-1} is the inverse operator to L. Note that whereas L is a linear differential operator, L^{-1} is a linear integral operator.

Now suppose that an observable data, d_i, is related to the scalar field, u(t), by an inner product:

$$d_i = (h_i(t), u(t)) \qquad (12)$$

This could be expressed in words as "the data, d_i, is a weighted average of the values of the scalar field, u(t), where the weighting is given by the function $h_i(t)$". The data kernel, $g_i(t)$, relates a perturbation, $\delta f(t)$, in forcing to a perturbation, δd_i, in the observed data:

$$\delta d_i = (g_i(t), \delta f(t)) \qquad (13)$$

The can now flesh out the original example of an egg cooking on a frying pan on a stove with a flickering flame. The flame is the source of heat, h(t), that governs the temperature, u(t) of the frying pan though a differential equation that describes heat transfer. The opacity, d_1, of the egg white after 3 minutes of cooking is the observable. It depends upon the temperature history of the frying pan, through the weighting function, $h_1(t)$. We might ask, how much is the opacity perturbed, if we were to increase the heat of the flame by 10% during the first minute of cooking? The data kernel, $g_1(t)$, provides the link between the perturbation in heat and the perturbation in opacity that allows this question to be answered. In subsequent discussions of the egg problem, we will omit the

subscript from the datum, d, and its corresponding weighing function, h(t), and data kernel, g(t), since in this problem there is only one observable.

3. A formula for the data kernel. The issue at hand is how to calculate the data kernel, $g_i(t)$, when L and $h_i(t)$ are known. We begin by combining equations 12 and 13 and then employing adjuncts:

$$d_i = (h_i, u) = (h_i, L^{-1}f) = ((L^{-1})^* h_i, f) = ((L^*)^{-1} h_i, f) \qquad (14)$$

Here we have used the identity, $(L^{-1})^* = (L^*)^{-1}$. Since the system is linear, this relationship also holds for perturbations, $\delta d_i = ((L^*)^{-1} h_i, \delta f)$, and thus from comparison with the definition of the data kernel in equation 9, we find that:

$$g_i(t) = (L^*)^{-1} h_i \quad \text{which is the solution to } L^* g_i = h_i \qquad (15)$$

In most applications, the scalar field, u(t), must be constructed by numerically solving the differential equation, $L\,u(t) = f(t)$ (equation 7). Thus the computational machinery is typically already in place to solve the companion differential equation, $L^* g_i = h_i$, and thus to construct the data kernel.

4. Causality. Many important physical systems are causal, meaning that for some specific time, τ, $u(\tau)$ is only influenced by values of f(t) for *past* times, $t \le \tau$. The matrix approximation, **C**, to the linear operator, L, is thus lower-triangular, since $u(t_i)$ can depend only upon $f(t_j)$ with $j \le i$. Writing $u_i = u(t_i)$ and $v_i = f(t_i)$, the differential equation Lu=f becomes **Cu=v**, or:

$$\begin{bmatrix} C_{11} & 0 & 0 & \dots & 0 \\ C_{21} & C_{21} & 0 & \dots & 0 \\ C_{31} & C_{32} & C_{33} & \dots & 0 \\ & & \dots & & \\ C_{N1} & C_{N2} & C_{N3} & \dots & C_{NN} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \dots \\ u_N \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \dots \\ v_N \end{bmatrix} \qquad (16)$$

Note that all the matrix elements above the main diagonal are zero, so **C** is *lower-triangular*. This matrix equation is solved by *forward-substitution*, since when read from top to bottom, each row (or equation) adds only one new unknown. The top equation involves only one unknown, u_1, and yields $u_1 = v_1/C_{11}$. The next equation involves two unknowns, u_1 and u_2, but the value of one of them is already known, so $u_2 = (v_2 - C_{21}u_1)/C_{22}$. And so forth. This procedure is equivalent to integrating the original differential equation *forward in time*. The matrix approximation to the linear operator, L^*, is just $\mathbf{C}^T$. It is upper-triangular, and can be solved by *back-substitution* (that is, starting with the last row and working upward). This procedure is equivalent to integrating the differential equation for the data kernel, $L^* g_i = h_i$, *backward in time*.

5. Finishing the Egg Problem. Suppose that the temperature, u(t), of the frying pan satisfies a simple Newton-style heat flow equation, du(t)/dt+ct=h(t) with the function h(t) quantifying the heat of the flame. This equation implies that in the absence of heating, the temperature of the pan decays away toward zero at a rate determined by the constant, c. We assume that the heating is restricted to some finite time interval near time zero, so that temperature satisfies the boundary conditions are $u(t\rightarrow\pm\infty)=0$.

First we solve the heat flow equation. The Green Function, which represents the response of the pan to an impulse of heat at time, t', solves the equation $dF(t,t')/dt + cu = \delta(t-t')$. It is:

$$F(t,t')=H(t-t')\exp\{-c(t-t')\} \qquad (17)$$

(where H(t–t'), a step function, is zero when t<t' and unity when t>t'). This result can be verified by first noting that equation 17 satisfies the boundary conditions and then by checking, though direct differentiation, that it solves the differential equation:

$$\begin{aligned} &L\,F(t,t') = \\ &\quad dF(t,t')/dt + cF(t,t') = \\ &\quad (d/dt)\,[H(t-t')\exp\{-c(t-t')\}] + cH(t-t')\exp\{-c(t-t')\} = \\ &\quad \delta(t-t')\exp\{-c(t-t')\} - cH(t-t')\exp\{-c(t-t')\} + c[H(t-t')\exp\{-c(t-t')\}] = \\ &\quad \delta(t-t')\exp\{-c(t-t')\} = \\ &\quad \delta(t-t') \qquad (18) \end{aligned}$$

The last step is valid since the value of $\exp\{-c(t-t')\}$ is unity at the point where the Delta function is non-zero. The temperature of the pan is zero before the heat pulse is applied, jumps to a maximum at the moment of application, and then exponentially decays back to zero at a rate determined bythe constant, c (figure 1).

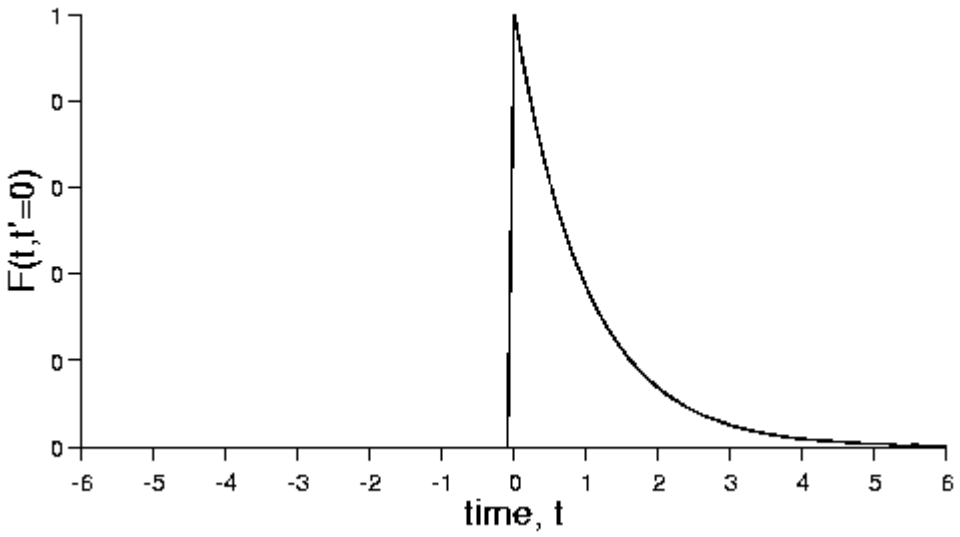

Figure 1. Green Function, F(t,t'), for the choice c=1 and t'=0.

To proceed further, we will need to specify a specific functional form for the averaging function, h(t). We will assume that it is simply a boxcar function, S(t,0,3), that is unity in the interval (0,3) and zero outside this interval. This choice implies that the opacity of the egg white integrates up the temperature of the frying pan between the times that it is put in the pan and taken off. The adjoint to the differential operator in equation 12 is $L^*=(-d/dt+c)$, since, as noted in Section 1, the adjoint of d/dt is (−d/dt) and the constant, c, is self-adjoint.

We now solve the adjoint equation. The Green Function of the adjoint equation solves $L^*G(t,t')=\delta(t-t')$, with $L^*=(-d/dt+c)$, and is:

$$G(t-t')=\{1-H(t-t')\}\exp\{c(t-t')\} \tag{19}$$

This result can be verified by first noting equation 19 satisfies zero boundary conditions at ±∞ and by demonstrating, though direct differentiation, that it solves the differential equation:

L^* G(t,t') =

dG(t,t')/dt + cG(t,t') =

(−d/dt) [{1−H(t−t')}exp{c(t−t')}] + c{1−H(t−t')}exp{−c(t−t')} =

δ(t−t')exp{c(t−t')} − c{1−H(t−t')}exp{c(t−t')} + c[{1−H(t−t')}exp{c(t−t')}] =

δ(t−t') exp{−c(t−t')} = δ(t−t') (20)

Thus the solution of the data kernel equation, $L^*g=h$, (equation 15) is:

$$g(t) = \int_{-\infty}^{+\infty} G(t,t')\, h(t')\, dt' = \int_{-\infty}^{+\infty} G(t,t')\, S(t',0,3)\, dt' = \int_{0}^{3} G(t,t')\, dt' = \exp(ct) \int_{0}^{3} \{1-H(t-t')\}\exp\{-ct'\}dt' \quad (21)$$

Note that the step function, $\{1-H(t-t')\}$, is unity when $t<t'$ and zero when $t>t'$. In order to perform the integration in equation 21, we must consider three cases, where time, t, is to the left of, within, and to the right of the interval of integration:

Case $t<0$. For times, t, less than zero, the step function is unity everywhere in the interval $0<t'<3$, since t' always greater than t.

$$g(t) = \exp(ct) \int_{0}^{3} \exp(-ct')\, dt' = (1/c)\, [1 - \exp(-3c)]\exp(ct) \quad (22a)$$

Case $0 \le t \le 3$. For times, t, in the interval [0,3], the step function starts out as zero, since $t'<t$, and then jumps to unity when $t'=t$. Thus lower limit of integration can be changed to t.

$$g(t) = \exp(ct) \int_{t}^{3} \exp(-ct')\, dt' = (1/c)\, [\exp(-ct) - \exp(-3c)]\, \exp(ct) \quad (22b)$$

Case $t>3$. For times, t, greater than 3, the step function is zero everywhere in the interval $0<t'<3$, since t' always less than t. Thus the integral is zero.

$$g(t)=0 \quad (22c)$$

We plot the data kernel, g(t), for $c=1$ (figure 2). Note that it is continuous across the points $t=0$ and $t=3$.

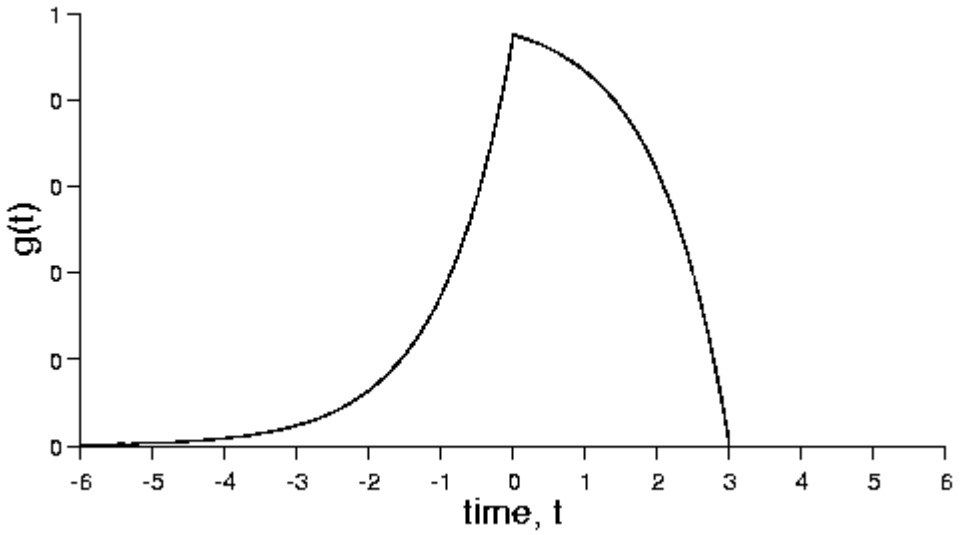

Figure 2. The data kernel g(t) for the choice c=1.

Suppose that the perturbations, δf, in heating is limited to a short time interval centered at time, t', so that it can be approximated by the Dirac function $\delta(t-t')$. The perturbations in egg white opacity, δd is given by the inner product $\delta d=(g,\delta f)$ (equation 13), which is just $\delta d=(g,\delta f)=(g,\delta(t-t'))=g(t')$.

For heating perturbations that occur before the egg is put on the pan, that is before t=0, $\delta d \propto \exp(ct')$ (equation 22a). Perturbations in heating occurring well before time t=0 when the egg is put on the pan have exponentially less impact on opacity than perturbations that occur just a little prior to that time. This behavior represents the effect of the cooling of the pan, which damps out the effect of heating perturbations in the far past.

Heating perturbations that occur after then egg is taken off the pan at time t=3 have no impact on egg white opacity. Thus, g(t) is zero for times, t>3.

Perturbations in heating that occur during the interval $0 \le t \le 3$ have the largest impact when they occur at the beginning of the interval and little impact when they occur at the end of the interval. This behavior can be understood from figure 1, which shows that a heating perturbation leads to a long-lasting perturbation in pan temperature. Heating perturbations that occur at the beginning of the cooking interval have an impact on pan temperature throughout the interval.

The original question that was posed was, "How much is the opacity perturbed, if we were to increase the heat of the flame by 10% during the first minute of cooking"? The perturbation in heating is thus $\delta f(t)=0.1f(t)S(t,0,1)$. Equation 13 relates the perturbation in

heating, δh, to the perturbation in opacity, δd, through the inner product, $\delta d=(g(t),\delta f)$. Since the heating perturbation is non-zero only during the cooking interval $0\le t\le 3$, we can use the form of the data kernel in equation 22b and write:

$$\delta d = (g(t),\delta f) = \int_0^1 (1/c)\,[\exp(-ct) - \exp(-3c)]\exp(ct)\,0.1\,f(t)\,dt \qquad (23)$$

In order to perform the integration, and thus to achieve an analytic expression for the perturbation in opacity, we would need to specify the actual form of the heating function, f(t). Equation 23 is easily integrable in the special case of a constant heating function, f(t)=1, but we leave it to the reader to work out the details.

7. An aside involving convolutions. In the example above, we were concerned with the effect of a perturbation in forcing on the data. Another common problem is to understand the effect of a perturbation of a parameter in the differential operator, *L*, on the data. The parameter, c, in the differential operator, $L=d/dt+c$, would be an example.

This problem is often approached using a Born-type approximation that converts the perturbation, δc, in the parameter, c, into an equivalent perturbation, δf, in forcing, f. It works like this: suppose that u(t) solves the original equation, $[d/dt+c]\,u(t) = f(t)$, and that $u(t)+\delta u(t)$ solves the equation with the perturbed parameter, $[d/dt+c+\delta c]\,[u(t)+\delta u(t)] = f(t)$, or:

$$du/dt + cu + \delta c\,u + d\delta u/dt + c\delta u + \delta c\,\delta u = f(t) \qquad (24)$$

We now subtract out the unperturbed equation, ignore the $\delta c\,\delta u$ term (presuming the product of two perturbations to be negligibly small) and rearrange:

$$d\delta u/dt + c\delta u = -\delta c\,u \qquad (25)$$

Thus the perturbation, δc, in the parameter, c, is equivalent to a perturbation, $df=-u\delta c$, in force, f. We can then use the formalism that we developed above to determine the effect of this perturbation on the data.

An important special case is when the data is the perturbed field, $\delta u(t)$, itself, that is, $d_i=\delta u(\tau_i)$, where τ_i is some specific time. This choice implies that that the weighting function in equation 12 is a Dirac delta function, $h_i(t)=\delta(t-\tau_i)$, and that the data kernel is the solution to the adjoint equation, $L^*\,g_i = \delta(t-\tau_i)$, that is the data kernel g_i is the Green's function of L^*, which we will denote $G(t,\tau_i)$. By equation 13, the perturbation in data is:

$$d_i = \delta u(\tau_i) = (\,g_i(t),\,\delta f(t)\,) = -\delta c \int_{-\infty}^{+\infty} G(t,\tau)\,u(t)\,dt \qquad (26)$$

In this example, none of the parameters in the adjoint differential equation are explicitly functions of time, and so the equation has *translational invariance*. Its Green's function can depend only upon the time difference, $t-\tau$, and not upon the actual value of either t or τ. The Green's function must therefore have the form, $G(t,\tau)=G(t-\tau)$. Equation 26 thus contains a convolution of the Green's function and the solution of the original, unperturbed, differential equation:

$$\delta u(\tau_i) = -\delta c \int_{-\infty}^{+\infty} G(t-\tau)\, u(t)\, dt \qquad (26)$$

7. A note on inversion. In this tutorial, we have used the fundamental equation for the data kernel, $\delta d_i = (g_i(t),\delta f(t))$, as a way of assessing the impact of a perturbation of forcing, $\delta f(t)$, on the perturbation of observed datum, δd_i. In many practical application, it's the inverse assessment that is important: we start out with a guess for the forcing, f(t), that imperfectly predicts that data, d_i, and then want to calculate a corrective perturbation to that forcing, $\delta f(t)$, that brings the predicted data closer to the observed data. The body of techniques that enable the forcing, f(t), to be reconstructed using many, possibly noisy, observations d_i, is called *inverse theory*. The data kernel, $g_i(t)$, plays a central role in inverse theory.